How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines for using clinical and economic evaluations

Andreas Laupacis, MSc, MD, FRCPC; David Feeny, PhD; Allan S. Detsky, MD, PhD, FRCPC; Peter X. Tugwell, MSc, MD, FRCPC

Because economic evaluations of health care services are being published with increasing frequency it is important to (a) evaluate them rigorously and (b) compare the net benefit of the application of one technology with that of others. Four "levels of evidence" that rate economic evaluations on the basis of their methodologic rigour are proposed. They are based on the quality of the methods used to estimate clinical effectiveness, quality of life and costs. With the use of the magnitude of the incremental net benefit of a technology, therapies can also be classified into five "grades of recommendation." A grade A technology is both more effective and cheaper than the existing one, whereas a grade E technology is less or equally effective and more costly. Those of grades B through D are more effective and more costly. A grade B technology costs less than $20 000 per quality-adjusted life-year (QALY), a grade C one $20 000 to $100 000/QALY and a grade D one more than $100 000/QALY. Many issues other than cost effectiveness, such as ethical and political considerations, affect the implementation of a new technology. However, it is hoped that these guidelines will provide a framework with which to interpret economic evaluations and to identify additional information that will be useful in making sound decisions on the adoption and utilization of health care services.

Puisqu'on publie de plus en plus souvent des évaluations économiques des services de soins de santé, il est important (a) d'évaluer rigoureusement ces derniers et (b) de comparer l'avantage net de l'utilisation d'une technologie par rapport à d'autres. Quatre «niveaux factuels» sont proposés pour coter les évaluations économiques d'après leur rigueur méthodologique. Ces niveaux reposent sur la qualité des modèles utilisés pour évaluer l'efficacité clinique, la qualité de la vie et les coûts. En tenant compte de l'ampleur des avantages cumulatifs nets d'une technologie, on peut également classer les thérapies en cinq «cotes de recommandation». Une technologie de cote A est à la fois efficace et moins coûteuse que la technologie en place, tandis qu'une technologie de cote E est tout au plus aussi efficace, mais plus coûteuse. Les technologies de cotes B à D sont plus efficaces et plus coûteuses. Une technologie de cote B coûte moins de 20 000 $ par année de vie pondérée par la qualité (AVPQ), une technologie de cote C, 20 000 $ à

Dr. Laupacis is an associate professor in the departments of Medicine and of Epidemiology and Community Medicine, University of Ottawa, Ottawa, Ont.; Dr. Feeny is professor in the departments of Economics and of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ont.; Dr. Detsky is professor in the departments of Health Administration and Medicine, University of Toronto, Toronto, Ont.; and Dr. Tugwell is professor in the departments of Medicine and of Epidemiology and Community Medicine, University of Ottawa, Ottawa, Ont.

Reprint requests to: Dr. Andreas Laupacis, Clinical Epidemiology Unit, Loeb Research Institute, Ottawa Civic Hospital, 1053 Carling Ave., Ottawa, ON K1Y 4E9

← For prescribing information see page 632

100 000 $ par AVPQ et une cote D, plus de 100 000 $ par AVPQ. En plus de la rentabilité, nombre de questions, par exemple les aspects déontologiques et politiques, influent sur la mise en application d'une nouvelle technologie. Cependant, on espère que les présentes lignes directrices offriront un cadre pour l'interprétation des évaluations économiques et l'identification de l'information supplémentaire qui sera utile pour prendre de bonnes décisions dans l'adoption et l'utilisation des services de soins de santé.

The number of published studies that include economic evaluations of health care services has increased in recent years, spurred by the large number of new therapeutic and diagnostic technologies, their associated costs and the limited resources available to pay for them. For the results of clinical and economic evaluations to be used for policy formulation it is important to develop an idea of the orders of magnitude of cost-effectiveness that are likely to be associated with wise adoption and utilization and with unwise use of health care resources.

How clinically and economically attractive does a technology have to be to warrant adoption and utilization? Although there is no definitive answer to the question, we propose a classification system in this article that provides guidance on the use of clinical and economic evaluations in making decisions about the adoption and utilization of competing health care technologies. Examples are provided of how published studies would be categorized, and the potential uses and limits of the system are discussed.

With the proposed system it will be possible to summarize the results of clinical and economic evaluations of health care technologies in terms of both the methodologic quality of the evaluations (levels of evidence) and the likely magnitude of net benefit from their application (grades of recommendation). The proposed classification scheme is modelled after the work of the Canadian Task Force on the Periodic Health Examination[1] and the National Institutes of Health–American College of Chest Physicians Task Force on the Use of Anti-thrombotic Agents.[2]

Levels of evidence

A complete economic evaluation considers both the effectiveness and the costs and includes the following six items (as adapted from reference 3).

1. All relevant clinical outcomes and costs are included in the analysis and valued sensibly. It is important to consider the methods used to establish effectiveness, estimate quality of life and measure costs. Criteria with which to assess the quality of these methods are provided in Appendix 1.

2. The analysis is incremental in that it compares the differences in costs and clinical outcomes of one specific technology (or policy) with those of another.

3. Costs and clinical outcomes are discounted.

4. Sensitivity analyses are used to assess the robustness of the conclusions.

5. The perspective of the decision-maker is clearly identified. This is usually the societal perspective, although it may be appropriate to take a purely organizational perspective (e.g., the hospital's) if the economic attractiveness of various options is being ranked within that organization.

6. The incremental cost–utility ratio identified must be compared with others in order to determine the economic attractiveness of one program over that of another.

A full economic study includes all six items. The level of evidence provided by such a study depends on the methodologic quality of the assessment of effectiveness, quality of life and costs (Appendix 1). A level I study uses the highest-quality assessment method for each of these three components, a level II study uses the highest-quality method for two, and a level III study uses the highest-quality method for one. All other studies are classified as level IV.

An example of a level I study is the economic comparison of a community-based treatment program for chronically disabled psychiatric patients and in-hospital management[4,5] (this was actually a cost–benefit analysis, so improvements in outcome were translated into dollar values rather than expressed as quality-adjusted life-years [QALYs]). Patients were randomly allocated to either type of care, resource use was collected prospectively, and costs were appropriately valued.

An example of a level II study is the economic evaluation of neonatal intensive care units by Boyle and associates.[6] Although the estimation of costs and quality of life was of high quality, effectiveness was assessed with a before–after study design.

It is recognized in some instances that the effectiveness of an intervention is so dramatic that a randomized controlled trial is not possible (e.g., heart transplantation v. no transplantation in patients with end-stage heart failure). In other instances the logistics of performing a randomized controlled trial are virtually insurmountable because the outcome of interest is so rare (e.g., evaluating universal precautions to prevent the spread of human immunodeficiency virus [HIV] infection).

Nevertheless, it is hoped that classifying studies into levels I through IV will allow the reader to be aware of the quality of the evidence.

Grades of recommendation

The decision about whether to implement a new therapy depends not only on the levels of evidence (the quality of the study) but also on the likely magnitude of the incremental costs required to achieve each additional unit of benefit. The suggested grades of recommendation (Table 1) classify therapies on the basis of the magnitude of their incremental net benefits.

A grade A technology is both more effective and less costly than the existing technology. There are, therefore, compelling reasons to introduce it or use it appropriately. Although most health care technologies do not meet the criteria for a grade A recommendation screening for phenylketonuria[7] and postpartum anti-D therapy[8] are examples that do.

Grade B through D technologies are classified as those that are (a) more effective and more costly than the existing technology or (b) less effective and less costly. Whether technologies are classified as grade B, C or D depends on the magnitude of the change in costs relative to outcome associated with their introduction (less than $20 000/QALY, $20 000 to $100 000/QALY or more than $100 000/QALY).

In this classification changes are measured relative to the costs and effects of the current policy, and a technology is classified as (a) or (b) depending on whether it has already been introduced into the health care system. For example, it has been estimated that the introduction of universal precautions to prevent HIV transmission to health care workers costs about $565 000 per additional life-year saved.[9] If universal precautions had not yet been introduced they would have been classified as grade Da (weak evidence for adoption or appropriate utilization). However, if a health care jurisdiction has already introduced universal precautions their abandonment is classified as grade Bb (in this case the "new" technology is standard precautions, and its abandonment would save more than $100 000/QALY). This is illustrated in Fig. 1.

In general, it seems harder to withdraw an expensive and relatively ineffective technology than to introduce an equally expensive and more effective one. However, some health care technologies have been adopted on the basis of weak clinical evidence

Table 1: Grades of recommendation for the adoption and appropriate utilization of new technologies

A. Compelling evidence for adoption and appropriate utilization
   The new technology is as effective as or more effective than the existing one and is less costly.
B. Strong evidence for adoption and appropriate utilization
   a) The new technology is more effective than the existing one and costs less than $20 000 per quality-adjusted life-year (QALY) gained.
   b) The new technology is less effective than the existing one, but its introduction would save more than $100 000/QALY gained.
C. Moderate evidence for adoption and appropriate utilization
   a) The new technology is more effective than the existing one and costs $20 000 to $100 000/QALY gained.
   b) The new technology is less effective than the existing one, but its introduction would save $20 000 to $100 000/QALY gained.
D. Weak evidence for adoption and appropriate utilization
   a) The new technology is more effective than the existing one and costs more than $100 000/QALY gained.
   b) The new technology is less effective than the existing one, but its introduction would save less than $20 000/QALY gained.
E. Compelling evidence for rejection
   The new technology is less effective than or as effective as the existing one and is more costly.

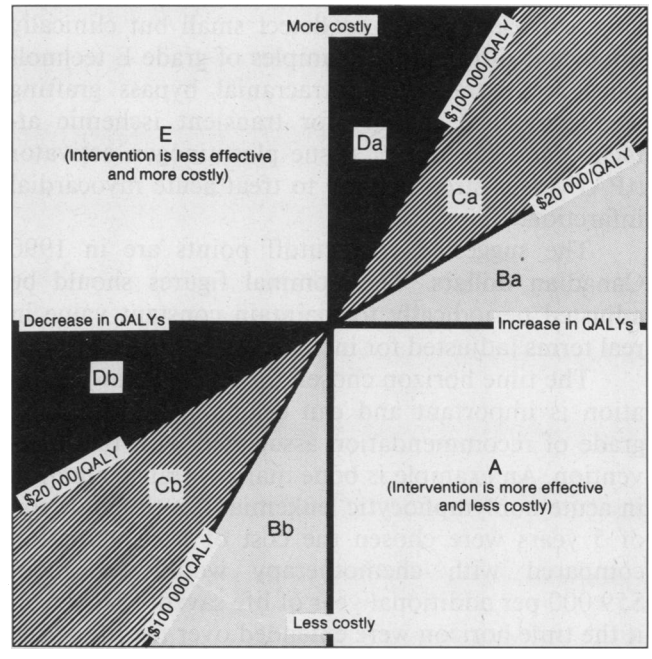

Fig. 1: Grades of recommendation: Grade A technologies should clearly be introduced or continued, and grade E technologies should not be introduced or should be abandoned. Technologies in the upper right quadrant are more effective and more costly than their alternatives, whereas those in the lower left quadrant are less effective and less costly. Introduction of technologies in the upper right quadrant and abandonment of the technologies with the same shading in the lower left quadrant lead to similar degrees of cost-effectiveness.

of effectiveness and without any formal economic evaluation. Thus, if resources can be saved (and put to better use elsewhere) policies that are relatively cost ineffective should be abandoned. On the other hand, one could argue that because standard practice has been in place for some time one should require a strong argument to justify a change.

For the sake of clarity the examples we will give of grades B through D technologies will be confined to those that are both more effective and more costly. Grade Ba technologies include coronary artery bypass grafting for left main coronary artery disease,[10] neonatal intensive care for infants weighing 1000 to 1499 g[6] and treatment in men with a diastolic blood pressure of 105 mm Hg or more.[11] An example of a grade Ca technology is hospital hemodialysis,[12] and examples of Grade Da technologies are the treatment of asymptomatic hyperlipidemia with cholestyramine,[13] the use of nonionic contrast media in patients at low risk of side effects[14] and the management of patients with low-risk myocardial infarction in a coronary care unit instead of intermediate care.[15]

Grade E technologies are more costly than existing technologies and less (or equally) effective. Before one can conclude that two technologies are equally effective the studies evaluating them must have sufficient power to detect small but clinically important differences. Examples of grade E technologies are extracranial–intracranial bypass grafting versus medical therapy for transient ischemic attacks[16] and the use of tissue plasminogen activator (tPA) versus streptokinase to treat acute myocardial infarction.[17]

The suggested cost cutoff points are in 1990 Canadian dollars. The nominal figures should be adjusted periodically to maintain constant value in real terms (adjusted for increases in the price level).

The time horizon chosen for an economic evaluation is important and can dramatically affect the grade of recommendation associated with the intervention. An example is bone marrow transplantation in acute nonlymphocytic leukemia. If a time horizon of 5 years were chosen the cost of transplantation compared with chemotherapy would be about $59 000 per additional year of life saved.[18] However, if the time horizon were extended over the life of the patient, then the cost would be about $10 000 per additional year of life saved.

Combining levels of evidence and grades of recommendation

The levels and grades can be combined to provide a summary of both the methodologic quality of the evidence and the magnitude of the net benefit associated with the therapy. For example, neonatal intensive care for infants weighing 1000 to 1499 g[6] is a Ba-II technology.

In some instances the evidence of either effectiveness or costs provided by methodologically suboptimal studies (e.g., those of levels II through IV) may be sufficient to justify its use in decision-making. Imagine an extremely expensive technology for which the clinical evidence is weak (e.g., no evidence from randomized controlled trials). One can take the most extreme assumption in favour of the intervention, and if the technology is economically unattractive under these conditions, then one is quite certain that it will still be unattractive even if a higher-quality method of assessing its effectiveness is used.

Such a situation arose recently with the introduction of tPA for the treatment of acute myocardial infarction. Despite the lack of a randomized controlled study comparing the rates of death among patients receiving tPA or streptokinase the greater cost of tPA (10 times that of streptokinase) was sufficient to persuade both the Ontario Medical Association and the Ontario government not to provide hospitals with special funding for tPA until evidence supporting its superiority over streptokinase was forthcoming.[19] This decision was made in 1988, and in 1990 the results of a direct comparison of the two agents showed that tPA was not more effective.[17]

Choice of cost/QALY cutoff limits

The grades of recommendation divide technologies into those that cost (or save) less than $20 000/QALY, $20 000 to $100 000/QALY or more than $100 000/QALY. These arbitrary limits were chosen after a review of available economic evaluations and previously suggested guidelines.[20] Technologies that cost less than $20 000/QALY are almost universally accepted as being appropriate ways of using society's and the health care system's resources. Many technologies costing $20 000 to $100 000/QALY are provided routinely, but the availability of some is significantly limited (e.g., elective coronary artery bypass grafting[21]), and there is discussion about the appropriateness of others for various patient groups (e.g., bone marrow transplantation for those over 45 years of age[18]).

Two technologies can be classified in the same level and yet be very different in cost effectiveness. The administration of nonionic contrast media to people at high risk[14] ($23 000/QALY) and hospital hemodialysis[12] ($65 500/QALY) are both grade Ca technologies. However, the techniques of economic evaluation and quality-of-life assessment are not as yet standardized. The calculated cost/QALY can vary considerably depending on the techniques used. Also, reasonable sensitivity analyses may change the

cost-effectiveness of an intervention greatly. Thus, we felt that narrowing the cost/QALY ranges of the various levels any further was not justified on the basis of currently available empirical evidence and analytical techniques.

Choice of clinical outcomes

QALYs have been suggested as an appropriate outcome measure for economic evaluations because they provide a "common yardstick" with which to compare the effectiveness of various interventions. QALYs are an index, a composite of the extra years of life provided by a therapy and the quality of that life, as measured by utilities.[22] By convention the utility scale runs from 0 to 1, 0 being equivalent to indifference between life and death and 1 being perfect health. Utilities can be estimated empirically by interviewing the investigators, the health care workers, members of society or the patients. The two most frequently used methods of measuring utilities in patients are the standard gamble and time trade-off techniques, although utilities can also be derived from multiattribute health indexes.[6,23] In general, measurements of patient or societal preferences are preferred for assessing health care technologies and forming policy.

Some limitations of utilities have recently been pointed out:[24-27] techniques are not standardized for measurement (this may yield different results in the same group of patients), utilities may be relatively unresponsive to a clinically important change detected by other outcome measures, and QALYs may not always accurately reflect the preferences of patients. Despite these reservations QALYs still seem to be a reasonable outcome measure for use in economic evaluations.

Some methodologically sound evaluations may not use QALYs; instead they describe the outcomes as discrete clinical events (e.g., myocardial infarction prevented or gastrointestinal hemorrhage avoided). It is unclear how such studies should be incorporated into the proposed grades of recommendation. At present, provided the outcome prevented is of major clinical importance, we suggest that an estimate of the utility associated with each event prevented (derived from asking either experts or patients) be used to calculate QALYs. However, it should be clearly indicated that QALYs were not assessed in the original study. As investigators gain more experience with QALYs it may become apparent that the utility associated with a clinical outcome (e.g., myocardial infarction prevented) is similar in different populations. If so, utilities would not have to be measured in all economic evaluations.

There are some interventions for which QALYs are difficult to measure. The calculation of cost-utility ratios for interventions that reduce short-term disabilities (e.g., the nausea, vomiting or pain associated with the use of contrast agents or postoperative recovery) is difficult, because these disabilities constitute such a small proportion of a person's entire life. Also, the utility derived from a reduction in uncertainty (e.g., the diagnosis of multiple sclerosis with the use of magnetic resonance imaging) cannot easily be incorporated into a full cost-utility analysis. An alternative approach is to determine which patients would be willing to pay for the reduction in disabilities or uncertainty. For example, in a sample of outpatients most were unwilling to pay $50 to decrease the risk of minor side effects from contrast media (pain, nausea, hives and flushing), but the median willingness to pay to reduce major and minor side effects from low-osmolar contrast media was $50.[28]

Incorporating the guidelines

The proposed guidelines offer direction concerning the strength of evidence for clinical and economic effectiveness associated with changes in health care policy. The guidelines are proposed as a necessary but not sufficient step in making decisions about the adoption and utilization of new technologies. The use of such guidelines would have a number of implications for both the conduct of clinical and economic evaluations and the forming of health care policy. These implications are briefly discussed below.

Timing of economic evaluations

The ideal time to evaluate the cost-effectiveness of a technology is before its widespread introduction into clinical practice, preferably at the same time as the randomized controlled trial is conducted to measure its clinical efficacy or shortly thereafter. This is rarely done. There are many reasons for this. Economic data are not required for the approval or licensing of most drugs and nonpharmaceutical technologies, and therefore there is no incentive for manufacturers to perform or encourage such evaluations (indeed, economic evaluations might indicate that the new technology is relatively cost-ineffective). Many physician researchers are interested in the clinical benefits of the technology but not the costs. Adding an economic evaluation to a clinical assessment can be expensive in terms of expertise, personnel and costs, and there is thus a reluctance to perform an economic evaluation before the clinical efficacy of the technology has been established.

However, as with many health care interventions, if the technology is found to be effective it is often incorporated into routine clinical practice be-

fore an economic evaluation can be done. This has occurred with recombinant human erythropoietin, which was recently licensed for use in patients with end-stage renal failure. The various provincial governments quite understandably felt obliged to decide on the level of funding for the drug before a complete economic evaluation was available, although they did have access to some economic evaluations funded by the pharmaceutical company. Lobbying from patients, nephrologists and the manufacturer made it impossible for the governments to delay their decision any longer, even though an economic evaluation was undertaken while the drug was being evaluated clinically.[29] However, the time required to perform the economic evaluation did not enable it to be peer-reviewed and published before the funding decision had to be made. In addition, some data needed for a complete economic evaluation will not be available for many years (e.g., employment status of recipients and the long-term cardiovascular effects of the drug).

Finally, the incremental cost-effectiveness of a technology at the point of its introduction may be very different from its cost-effectiveness later on, because as the technology gets better its incremental cost-effectiveness ratio improves (as is the case with liver transplantation).

Selection of technologies for economic evaluation

Like resources for the health care system itself, funds for evaluative studies are limited. A full-scale economic evaluation can add considerable cost to a clinical study, and it would be unfeasible to perform extensive economic evaluations on all new technologies. Such analyses are relatively unimportant when the condition is extremely rare and the total cost relatively minor. Economic evaluations should be performed if technologies are either extremely costly per case (e.g., bone marrow transplantation) or likely to be used by a considerable proportion of the population (e.g., nonionic contrast media) and are therefore potentially costly in aggregate.

To date, relatively few economic evaluations of diagnostic technologies (in terms of the equipment and the manner physicians use it) have been published. This is due in part to the difficulty in evaluating many of these technologies because they are used for a wide variety of indications, and the choice of the alternative diagnostic modality depends on the indication. However, studies are now becoming available that directly compare the diagnostic accuracy, sensitivity and specificity of different technologies. Recent examples include magnetic resonance imaging versus transrectal ultrasonography for the staging of clinically localized prostatic carcinoma[30] and magnetic resonance imaging versus computed tomography for patients with suspected lesions in the posterior cranial fossa.[31] Few of these studies provide accompanying economic evaluations, which should be encouraged in the future.

Given that many health care technologies are adopted and used in the absence of any evidence from systematic evaluation it is reasonable to question the usefulness of the proposed guidelines. It can be argued, however, that the early application of these guidelines could marginally improve the situation even when comprehensive evaluations are unavailable. The attempt to apply the guidelines will help to identify major gaps in the information on effectiveness, quality of life and costs. Even if it is impossible to organize a high-quality evaluative study to remediate the identified deficiencies, it should be possible to collect expert opinion systematically for each type of information and assemble it to assist in decision-making. The report would be classified as a level IV study, and the lack of strength of the evidence would then be included explicitly in the deliberations concerning the technology. The guidelines are important as both a means of grading evidence and a framework that identifies the types of information that would be useful in making sound decisions about adoption and utilization.

Total versus incremental costs

Our proposed classification system uses incremental cost–utility ratios rather than average ratios.[32] In an incremental analysis the differences in

Table 2: Introduction of low-osmolar contrast media as an example of an incremental cost–utility analysis

Treatment program	Cost per test, $	Utility (QALY)	Average cost–utility ratio,* $/QALY
Old	14.39	29.9986	0.48
New	36.98	29.9996	1.23

*The incremental cost–utility ratio for the new program was calculated by dividing the difference in the cost per test by the difference in the QALY ($22.59 ÷ 0.001 = $22 590/QALY).

both costs and consequences between new and old treatments are compared. This allows scarce resources to be allocated so that the maximum clinical benefit is provided.

In Table 2 the introduction of low-osmolar contrast media is used as an example of an incremental cost–utility analysis.[14] Two strategies are compared: the continued use of the old, high-osmolar media in all patients (the "old" program) and the use of the new, low-osmolar media only in patients at high risk of an adverse reaction (the "new" program). The average cost per patient of the contrast media as well as the average QALYs after contrast injection (assuming a life expectancy of 30 years with no adverse reaction) were calculated. The average cost–utility ratio of the new contrast media was $1.23/QALY ($36.98/29.9996). However, the incremental ratio was $22 600/QALY. Conceptually, the difference between these two ratios is that the incremental one reveals the cost per unit of the benefit of switching from one treatment strategy (usually already in use) to a new strategy, whereas the average ratio reflects the cost per benefit of the new strategy independent of alternative strategies. This example also illustrates that the old approach is not without cost — hence the need for an analysis of the costs and consequences of the old and new technologies.

However, decisions about and plans for the allocation of health care resources also consider the total costs of a technology. The number of patients undergoing long-term hemodialysis in a given year is less than the yearly incidence of myocardial infarction. Therefore, although two treatments may be about equal in terms of costs/QALY, the total cost for the treatment of myocardial infarction will be substantially greater than that for patients undergoing hemodialysis. The funding agency may be able to afford the latter but not the former. It will therefore be useful to include the number of patients who will benefit from the technology to assist in providing an estimate of the overall costs and benefits of the therapy.

Economic evaluations, ethics and politics

The introduction of a new technology is influenced by a combination of effectiveness, economics, ethics and politics. The relative contribution of each varies from situation to situation.

Economic evaluations deal with effectiveness and economics. However, society also needs to consider the ethical implications of health care policy when interpreting the results of a cost–benefit analysis.[33] For example, saving the life of a retired person may produce less direct economic benefits than saving the life of an employed person would.

The political process is the final pathway through which most decisions about the allocation of health care resources in Canada are made. Obviously factors other than effectiveness, economics and ethics come into play at this stage, and only a few will be briefly discussed here.

The perspective of an economic evaluation is extremely important. It is usually argued that a "societal" perspective, in which all costs and benefits associated with the introduction of a new program are considered, is the most appropriate. The ranking of cost-effectiveness ratios calculated from society's point of view should be neutral to value or distributional decisions. However, it is difficult for many people who decide on whether a program should be introduced to adopt an entirely societal point of view.[32] For example, the use of a cost-effectiveness analysis to forgo funding of a bone marrow transplant program will result in losses for patients with nonlymphocytic leukemia and gains for those who receive the alternatively funded interventions.

An institution may take its own point of view and rank cost-effectiveness ratios on which to base its allocation decisions. Alternatively some institutions may have particular goals that influence their resource allocation decisions independently of cost-effectiveness considerations. A hospital that sees itself as a tertiary care centre may wish to fund bone marrow transplantation rather than an immunization program. Physicians may stand to gain in financial terms and in terms of prestige if a program that they are associated with is funded. Thus, although a societal point of view is the most appropriate perspective many competing (and often legitimate) interests affect the allocation decision.

It is generally easier to withhold funding for a new technology than it is to withdraw funding from an existing one (even though the withdrawn funds could be spent more efficiently elsewhere). Now that universal precautions against HIV transmission have been introduced in some hospitals, it will be very difficult to withdraw them, even if they cost $565 000 per life saved. One example in which a more expensive and marginally safer technology has been withdrawn is the return to gentamicin as the aminoglycoside of choice in many hospitals.

Another influence on decision-making is the "identifiable beneficiary or victim." Programs that have an identifiable beneficiary or victim (e.g., a child with liver failure awaiting a transplant) often appear to receive higher priority than those that do not. Similarly, easily identifiable, "big-ticket" technologies (e.g., transplantation) receive much attention and discussion, whereas the frequently used and unnecessary "low-ticket" items (e.g., routine preoperative chest x-ray films in an asymptomatic patient)[34] may con-

sume more resources but receive little attention.

It is almost universally accepted that the funds available for health care are limited. However, the exact amount that Canada should spend is not at all clear. In 1986 Canada spent 8.5% of its gross domestic product (GDP) on health care.[35] This placed Canada third (along with France) among members of the Organization for Economic Co-operation and Development.[35] Only the United States and Sweden spent more of their GDP on health care (11.7% and 9.1% respectively). In a society as wealthy as ours it is clear that if more health care funding was a societal priority and if there was the political will, the available funds could still be increased. Nonmedical programs such as education could also benefit from more funding, and some of these programs affect health. However, the point is that although society's overall resources are limited the proportion that is spent on health care could be increased.

The guidelines proposed in this paper do not directly address the issue of determining how much in aggregate Canada should spend on health care. Their main purpose is to assist in deciding which technologies and programs should be funded within any given budget by focusing on evidence of their clinical and economic effectiveness. The applications of these guidelines could, however, assist in the making of broader policies concerning overall budget priorities. If, for instance, most technologies were classified as grade A or B the implication might be that health care would warrant an increase in the aggregate level of expenditures. On the other hand, if most were classified as grade D this would not be considered evidence to support an increase in the health care budget.

A final point concerns the medical profession itself. The economic evaluations discussed here may appear to have a very little role to play in the care provided by individual physicians. Patients go to their physicians expecting the best possible care, without consideration of costs. However, most physicians make economic decisions in their practices daily when they budget their time (spending more of it with patients whom they think they can help) and select tests or treatments (choosing the cheaper of equally useful ones). Also, in their administrative functions (as advisers to the government, medical chiefs of staff or heads of departments) physicians have a societal responsibility to ensure that the limited resources available for health care yield the maximum benefit. Many physicians have specialized practices and quite naturally find themselves acting as advocates for a subgroup of patients (nephrologists are much more likely to press for funding of erythropoietin than for the increased availability of hip replacements). If the government is to be persuaded to consider seriously the results of economic evaluations when allocating scarce resources, then physicians must encourage the conduct of such studies in all areas of medicine (not just those that support their own narrow interests) and be willing to be guided by the results.

We thank Dr. David Sackett for his helpful comments and Janice Coffey and Karen Weeks for preparing the manuscript.

Dr. Detsky was supported by a scholarship from the National Health Research and Development Program, Department of National Health and Welfare. Part of this work was done when Dr. Laupacis was a career scientist of the Ontario Ministry of Health.

References

1. Canadian Task Force on the Periodic Health Examination: The periodic health examination. *Can Med Assoc J* 1979; 121: 1193-1254
2. Sackett DL: Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest* 1986; 89: 25-35
3. Department of Clinical Epidemiology and Biostatistics, McMaster University Health Sciences Centre: How to read clinical journals: VII. To understand an economic evaluation (part B). *Can Med Assoc J* 1984; 130: 1542-1549
4. Stein LI, Test MA: Alternative to mental hospital treatment. *Arch Gen Psychiatry* 1980; 37: 392-397
5. Weisbrod BA, Test MA, Stein LI: Alternative to mental hospital treatment: 2. Economic benefit-cost analysis. Ibid: 400-405
6. Boyle MH, Torrance GW, Sinclair JC et al: Economic evaluation of neonatal intensive care of very low birth-weight infants. *N Engl J Med* 1983; 308: 1330-1337
7. Bush JW, Chen MM, Patrick DL: Health status index in cost-effectiveness: analysis of PKU program. In Berg RL (ed): *Health Status Indexes*, Hospital Research and Education Trust, Chicago, 1973: 172-209
8. Torrance GW, Zipursky A: Cost-effectiveness analysis of treatment with anti-D. Working paper presented at the Rh Prevention Conference, McMaster University, Hamilton, Ont, Sept 28-30, 1977
9. Stock SR, Gafni A, Bloch RF: Universal precautions to prevent HIV transmission to health care workers: an economic analysis. *Can Med Assoc J* 1990; 142: 937-946
10. Weinstein MC, Stason WB: Cost-effectiveness of coronary artery bypass surgery. *Circulation* 1982; 66 (suppl 3): 56-66
11. Stason WB, Weinstein MC: Allocation of resources to manage hypertension. *N Engl J Med* 1977; 296: 732-739
12. Churchill DN, Lemon DC, Torrance GW: Cost-effectiveness analysis comparing continuous ambulatory peritoneal dialysis to hospital hemodialysis. *Med Decis Making* 1984; 4: 489-500
13. Weinstein MC, Stason WB: Cost-effectiveness of interventions to prevent or treat coronary heart disease. *Annu Rev Public Health* 1985; 6: 41-63
14. Goel V, Deber RB, Detsky AS: Nonionic contrast media: economic analysis and health policy development. *Can Med Assoc J* 1989; 140: 389-395
15. Fineberg HV, Scadden D, Goldman L: Care of patients with a low probability of acute myocardial infarction: cost-effectiveness of alternatives to coronary-care-unit admission. *N Engl J Med* 1984; 310: 1301-1307
16. The EC/IC Bypass Study Group: Failure of extracranial-intracranial arterial bypass to reduce the risk of ischemic stroke: results of an international randomized trial. *N Engl J*

Med 1985; 313: 1191–1200
17. Fresco C, Franzosi MG, Maggioni AP et al: The GISSI-2 Trial: premises, results, epidemiological (and other) implications. *Clin Cardiol* 1990; 13: 32–36
18. Welch HG, Larson EB: Cost effectiveness of bone marrow transplantation in acute nonlymphocytic leukemia. *N Engl J Med* 1989; 321: 807–812
19. Linton AL, Naylor CD: Organized medicine and the assessment of technology. *N Engl J Med* 1990; 323: 1463–1467
20. Kaplan RM, Bush JW: Health-related quality of life measurement for evaluation research and policy analysis. *Health Psychol* 1982; 1: 61–80
21. Naylor CD, Baigrie RS, Goldman BS et al: Assessment of priority for coronary revascularisation procedures. Revascularisation Panel and Consensus Methods Group. *Lancet* 1990; 335: 1070–1073
22. Torrance GW, Feeny D: Utilities and quality-adjusted life years. *Int J Technol Assess Health Care* 1989; 5: 559–575
23. Boyle MH, Torrance GW: Developing multiattribute health indexes. *Med Care* 1984; 22: 1045–1057
24. Gafni A, Zylak CJ: Ionic versus nonionic contrast media: A burden or a bargain? *Can Med Assoc J* 1990; 143: 475–478
25. O'Connor AM, Boyd NF, Warde P et al: Eliciting preferences for alternative drug therapies in oncology: influence of treatment outcome descriptions, elicitation technique and treatment experience on preferences. *J Chronic Dis* 1987; 40: 811–818
26. Loomes G, McKenzie L: The use of QALYs in health care decision making. *Soc Sci Med* 1989; 28: 299–308
27. Mehrez A, Gafni A: Quality-adjusted life years, utility theory, and healthy-years equivalents. *Med Decis Making* 1989; 9: 142–149
28. Appel LJ, Steinberg EP, Power NR et al: Risk reduction from low osmolality contrast media. What do patients think it is worth? *Med Care* 1990; 28: 324–334
29. Sheingold SH, Churchill DN, Muirhead N et al: The impact of recombinant human erythropoietin on medical care costs for hemodialysis patients in Canada. *Soc Sci Med* (in press)
30. Rifkin MD, Zerhouni EA, Gatsonis CA et al: Comparison of magnetic resonance imaging and ultrasonography in staging early prostate cancer — results of a multi-institutional cooperative trial. *N Engl J Med* 1990; 323: 621–626
31. Teasdale GM, Hadley DM, Lawrence A et al: Comparison of magnetic resonance imaging and computed tomography in suspected lesions in the posterior cranial fossa. *BMJ* 1989; 299: 349–355
32. Detsky AS, Naglie IG: A clinician's guide to cost-effective analysis. *Ann Intern Med* 1990; 113: 147–154
33. Avorn J: Benefit and cost analysis in geriatric care. *N Engl J Med* 1984; 310: 1294–1301
34. Turnbull JM, Buck C: The value of preoperative screening investigations in otherwise healthy individuals. *Arch Intern Med* 1987; 147: 1101–1105
35. Schieber GJ, Poullier JP: International health spending and utilization trends. *Health Aff (Millwood)* 1988; 7 (4): 105–112
36. Sacks HS, Berrier J, Reitman D et al: Meta-analyses of randomized controlled trials. *N Engl J Med* 1987; 316: 450–455
37. Oxman AD, Guyatt GH: Guidelines for reading literature reviews. *Can Med Assoc J* 1988; 138: 697–703
38. Department of Clinical Epidemiology and Biostatistics, McMaster University Health Sciences Centre: How to read clinical journals: II. To learn about a diagnostic test. *Can Med Assoc J* 1981; 124: 703–710
39. Kirshner B, Guyatt G: A methodological framework for assessing health indices. *J Chronic Dis* 1985; 38: 27–36
40. Torrance GW: Measurement of health state utilities for economic appraisal. *J Health Econ* 1986; 5: 1–30
41. Guyatt GH, Veldhuyzen Van Zanten SJO, Feeny DH et al: Measuring quality of life in clinical trials: a taxonomy and review. *Can Med Assoc J* 1989; 140: 1441–1448
42. Boyle MH, Torrance GW, Harwood SP et al: *A Cost Analysis of Providing Neonatal Intensive Care to 500–1,499 gram Birth Weight Infants* (QSEP res rep 51), Program for Quantitative Studies in Economics and Population, Faculty of Social Sciences, McMaster University, Hamilton, Ont, 1982

Appendix 1: Criteria for assessing effectiveness, quality of life and costs in an economic evaluation

Effectiveness

Criteria for assessing the methodologic quality of studies that evaluate efficacy, meta-analyses and diagnostic tests have been published. The highest-quality evidence of efficacy is provided by a randomized controlled trial with low false-positive and false-negative error rates.[2] Alternatively, a meta-analysis in which only randomized controlled trials are included, a comprehensive search method is used to locate all the relevant studies, the variation in the findings of the studies are analysed, and the results of the primary studies are combined in an appropriate manner[36,37] can also provide the highest-quality evidence of efficacy. Some economic studies evaluate diagnostic tests. In such cases the highest-quality evidence is provided by a study in which there is an independent, blind comparison of two diagnostic tests, the sample of subjects with mild to severe disease is sufficiently large to allow narrow confidence limits on the resulting sensitivity, specificity or likelihood ratios, and the reproducibility of the test has been established.[38]

Quality of life

Quality of life should be estimated in an appropriate sample of people who have the disorder of interest with the use of measures whose validity, reliability and responsiveness to change have been demonstrated.[39] To allow comparison with cost–utility ratios for other interventions a utility measure (e.g., standard gamble, time trade-off and multiattribute utility measures) should be used that incorporates general health status rather than a disease-specific measure of quality of life.[40,41] However, if the economic attractiveness of various options in a particular program is being compared (e.g., treatment of end-stage renal disease) the use of disease-specific measures may be appropriate.

Costs

The highest-quality estimates of costs include those derived from direct measurement of resources used by the competing strategies in a sample of the population that used those resources. There are two components of costs: the volume of services used (e.g., hospital days, laboratory tests, physician visits) and the unit prices for each of those services. A high-quality study measures the volume of services and uses clearly identifiable unit prices that apply to the services (e.g., health ministry rates for physician services) and direct measurements of costs borne by institutions, including an appropriate method of allocating overhead.[42] Some investigators will have to use a sampling technique to estimate the quantity of individual services delivered under the competing strategies and use published sources for the unit prices.